32. Bioinformatics and Nanotechnologies: Nanomedicine

Federico Ambrogi, Danila Coradini, Niccolò Bassani, Patrizia Boracchi, Elia M. Biganzoli

In this chapter we focus on the bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers with a specific focus on breast cancer with a perspective on new diagnostic device tools coming from the field of nanobiotechnology and the challenges related to high-throughput data integration, analysis, and assessment from multiple sources.

Great progress in the development of molecular biology techniques has been seen since the discovery of the structure of deoxyribonucleic acid (DNA) and the implementation of a polymerase chain reaction (PCR) method. This started a new era of research on the structure of nucleic acids molecules, the development of new analytical tools, and DNA-based analyses that allowed the sequencing of the human genome, the completion of which has led to intensified efforts toward comprehensive analysis of mammalian cell struc-

ture and metabolism in order to better understand the mechanisms that regulate normal cell behavior and identify the gene alterations responsible for a broad spectrum of human diseases, such as cancer, diabetes, cardiovascular diseases, neurodegenerative disorders, and others.

Part F | 32

32.1 Background

Technical advances such as the development of molecular cloning, Sanger sequencing, PCR, oligonucleotide microarrays and more recently the development of a variety of so-called next-generation sequencing (NGS) platforms has actually revolutionized translational research and in particular cancer research. Now, scientists can obtain a genome-wide perspective of cancer gene expression useful to discover novel cancer biomarkers for more accurate diagnosis and prognosis, and monitoring of treatment effectiveness. Thus, for instance, microRNA expression signatures have been shown to provide a more accurate method of classifying cancer subtypes than transcriptome profiling and allow classification of different stages in tumor progression, actually opening the field of personalized medicine (in which disease detection, diagnosis, and therapy are tailored to each individual's molecular profile) and predictive

medicine (in which genetic and molecular information is used to predict disease development, progression, and clinical outcome).

However, since these novel tools generate a tremendous amount of data and since the number of laboratories generating microarray data is rapidly growing, new bioinformatics strategies that promote the maximum utilization of such data, as well as methods for integrating gene ontology annotations with microarray data to improve candidate biomarker selection are necessary. In particular, the management and analysis of NGS data requires the development of informatics tools able to assemble, map, and interpret huge quantities of relatively or extremely short nucleotide sequence data.

As a paradigmatic example, a major pathology such as breast cancer can be considered. Breast can-

cer is the most common malignancy in women with a cumulative lifetime risk of developing the disease as high as one in every eight women [32.1]. Several factors are associated with this cancer such as genetics, life style, menstrual and reproductive history, and long-term treatment with hormones. Until now breast cancer has been hypothesized to develop, following a progression model similar to that described for colon cancer [32.2, 3], through a linear histological progression from adenosis, to ductal/lobular hyperplasia, to atypical ductal/lobular hyperplasia, to in situ carcinoma and finally to invasive cancer, corresponding to increasingly worse patient outcome. Molecularly, it has been suggested that this process is accompanied by increasing alterations of the genes that encode for tumor suppressor proteins, nuclear transcription factors, cell cycle regulatory proteins, growth factors, and corresponding receptors, which provide a selective advantage for the outgrowth of mammary epithelial cell clones containing such mutations [32.4].

Recent advances in genomic technology have improved our understanding of the genetic events that parallel breast cancer development. In particular, DNA microarray-based technology, with the simultaneous evaluation of thousands of genes, has provided researchers with an opportunity to perform comprehensive molecular and genetic profiling of breast cancer able to classify it into some clinically relevant subtypes and in the attempt to predict the prognosis or the response to treatment [32.5–8]. Unfortunately, the initial enthusiasm for the application of such an approach was tempered by the publication of several studies reporting contradictory results on the analysis of the same samples analyzed on different microarray platforms that arose the skepticism regarding the reliability and the reproducibility of this technique [32.9, 10]. In fact, despite the great theoretical potential for improving breast cancer management, the actual performance of predictors, built using genes' expression, is not as good as initially published, and the lists of genes obtained from different studies are highly unstable, resulting in disparate signatures with little overlap in their constituent genes. In addition, the biological role of individual genes in a signature, the equivalence of several signatures, and their relation to conventional prognostic factors are still unclear [32.11]. Even more incomplete and confusing is the information obtained when molecular genetics was applied to premalignant lesions; indeed, genome analysis revealed an unexpected morphological complexity of breast cancer, very far from the hypothesized multi-step linear process, but sug-

gesting a series of stochastic genetic events leading to distinct and divergent pathways towards invasive breast cancer [32.12], the complexity of which limits the application of really effective strategies for prevention and early intervention.

Therefore, despite the great body of information about breast cancer biology, improving our knowledge about the puzzling bio-molecular features of neoplastic progression is of paramount importance to better identify the series of events that, in addition to genetic changes, are involved in breast tumor initiation and progression and that enable premalignant cells to reach the six biological endpoints that characterize malignant growth (self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis). To do that, instead of studying the single aspects of tumor biology, such as gene mutation or gene expression profiling, we must apply an investigational approach aimed to integrate the different aspects (molecular, cellular, and supracellular) of breast tumorigenesis.

At the molecular level, an increasing body of evidence suggests that gene expression alone is not sufficient to explain protein diversity and that epigenetic changes (i. e., heritable changes in gene expression that occur without changes in nucleotide sequences), such as alteration in DNA methylation, chromatin structure changes, and dysregulation of microRNA expression, may affect normal cells and predispose them to subsequent genetic changes with important repercussions in gene expression, protein synthesis, and ultimately cellular function [32.13–16]. At the cellular level, evidence indicates that to really understand cell behavior, we must consider also the microenvironment in which cells grow; an environment that recent findings indicate to have a relevant role in promoting and sustaining abnormal cell growth and tumorigenesis [32.17].

This picture is further complicated by the concept that among the heterogeneous cell population that makes up the tumor, there exists an approximately 1% of cells, also known as tumor initiating cells that are more likely derived from normal epithelial precursors (stem/precursor cells), and share with them a number of key properties including the capacity of self-renewal and the ability to proliferate and differentiate [32.18, 19]. When altered in their response to abnormal inputs from the local microenvironment, these stem/precursor cells can give rise to preneoplastic lesions [32.20]. In fact, similarly to bone marrow-derived stem cells, tissue-specific stem cells show remarkable

plasticity within the microenvironment: they can enter a state of quiescence for decades (cell dormancy), but can become highly dynamic once activated by specific microenvironment stimuli from the surrounding stroma and are ultimately transformed in tumor initiating cells [32.21]. The stroma, in which the mammary gland is embedded, is composed of adipocytes, fibroblasts, blood vessels, and an extracellular matrix in which several cytokines and growth factors are present. While none of these cells are themselves malignant, they may acquire an abnormal phenotype and altered function due to their direct or indirect interaction with epithelial stem/precursor cells. Acting as an oncogenic agent, the stroma could provoke tumorigenicity in adjacent epithelial cells leading to the acquisition of genomic changes, at which epigenetic alterations also concur, that can accumulate over time and provoke silencing of more than 100 pivotal genes' encoding for proteins involved in tumor suppression, apoptosis, cell cycle regulation, DNA repair, and signal transduction [32.22]. Under these conditions, epithelial cells and the stroma co-evolve towards a transformed phenotype following a process that has not yet been worked out [32.23, 24].

Many of the soluble factors present in the stroma, essential for the normal mammary gland development, have been found to be associated with cancer initiation. This is the case of hormone steroids (estradiol and progesterone), which are physiological regulators of breast development and whose dysregulation may result in preneoplastic and neoplastic lesions [32.25–27]. In fact, through their respective receptors, in epithelial cells estrogens and progesterone may induce the synthesis of local factors that, on the one hand, trigger the activation of the stem/precursor cells and, on the other hand, exert a paracrine effect on endothelial cells, which in response to the vascular endothelial growth factor, trigger neoangiogenesis activation [32.21]. In addition, estrogens have been found implicated in the local modifications of tissue homeostasis associated to a chronic inflammation that may promote epithelial transformation due to the continued production of pro-inflammatory factors that favors generation of a pro-growth environment and fosters cancer development [32.28]; alternatively, transformed epithelial cells would enhance activation of fibroblasts through a vicious circle that supports the hypothesis according to which cancer should be considered as a never healing wound. Last but not least, very recent findings in animal models have clearly indicated that an early event occurring in the activation of estrogen-induced mammary carcinogenesis is represented by the altered expression of some oncogenic microRNAs (oncomir), suggesting a functional link between hormone exposure and epigenomic control [32.29].

Concerning the forecasted role of new nanobiotechnology applications, disclosing the bio-molecular events contributing to tumor initiation is, therefore, of paramount importance and to achieve this goal a convergence of advanced biocomputing tools for cancer biomarker discovery and multiplexed nanoprobes for cancer biomarker profiling is crucial. This is the one of the major tasks currently ongoing in medical research, namely the interaction of nanodevices with cells and tissues in vivo and their delivery to disease sites.

## 32.2 Biostatistics Supporting Bioinformatics for Biomarker Discovery

Biomarkers refer to genes, RNA, proteins, and miRNA expressions that can be correlated with a biological condition or may be important for prognostic or predictive aims as far as regards the clinical outcome. The discovery of biomarkers has a long history in translational research. In more recent years, microarrays have generated a great deal of work, promising the discovery of prognostic and predictive biomarkers able to change medicine as was known until then. Since the beginning, the importance of statistical methods in such a context was evident, starting from the seminal paper of Golub, which showed the ability of gene expression to classify tumors [32.30].

Although bioinformatics is the leading engine, referenced in biomolecular literature, providing informatics tool to handle massive omic data, the computational core is actually represented by biostatistics methodology aiming at extracting useful summary information.

Biostatistics cornerstones are represented by large sample and likelihood theories, hypothesis testing, experimental design, and exploratory multivariate techniques summarized in the genomic era according to class comparison, prediction, and discovery.

Actually, massive omic data and the idea of personalized medicine need to develop statistical theory according to new requirements. Even in the case of multivariate techniques, the problems usually faced using statistical techniques accounted for orders of magnitude of less data than those encountered with
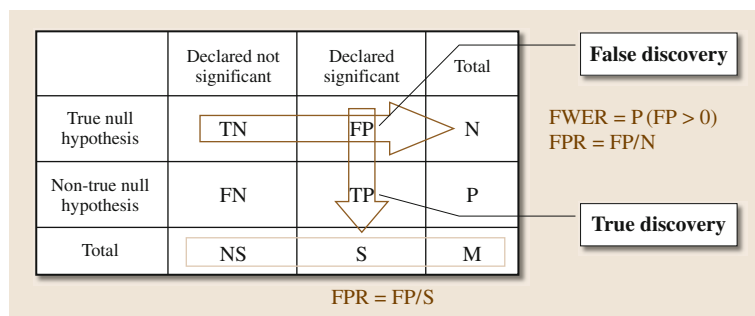
Fig. 32.1 Explanation of FDR (for further details, see [32.31])

high-throughput technologies. A situation that NGS techniques will easily exacerbate.

In class comparison studies there is a predefined group which identifies samples and the interest is in evaluating if the groups express the transcripts of interest differently. Such studies are generally performed using a transcript by transcript analysis, performing thousands of statistical tests then correcting p-values to account for the desired percentages of false positives and negatives. In fact, the multiple comparison problem is the first concern as traditionally methods for family-wise control are generally too restrictive when accounting for thousands of tests. The false discovery rate (FDR) was a major breakthrough in such a context. The general concepts underlying FDR are outlined later (Fig. 32.1).

Another topic discussed regards the parametric assumptions underlying most of the statistical tests used. Permutation tests were much developed to face this issue and are now one of the standard tools available to researchers.

Jeffery and colleagues [32.32] performed a systematic comparison of 9 different methods for identifying genes differentially expressed across experimental groups, finding that different methods gave rise to very different lists of genes and that sample size and noise level strongly affected predictive performance of the methods chosen for evaluation. Also, evaluation of the accuracy of fold-change compared to ordinary and moderated t statistics was performed by *Witten* and *Tibshirani* [32.33], which discusses the issues of reproducibility and accuracy of gene lists returned by different methods, claiming that

> *a researcher's decision to use fold-change or a modified t-statistic should be based on biological, rather than statistical, considerations.*

In this sense, the classical *limma-like* approach [32.34] has become a de facto standard in the analysis of high-throughput data: gene expression and miRNA

microarrays, proteomics, and serial analysis of gene expression (SAGE) generate an incredible amount of data which is routinely analyzed element-wise, without considering the multivariate nature of the problem. Akin to this, non-parametric multivariate analysis of variance (MANOVA) techniques have also been suggested to identify differentially expressed genes in the context of microarrays and qPCR-RT [32.35, 36], with the advantage of not making any distributional assumption on expression data and of being able to circumvent the dimensionality issue related to omic data ($n°$ of subjects $\ll n°$ of genes).

A well-known example of class comparison study was that of van't *Veer* and colleagues [32.37] in which a panel of genes, a signature, was claimed to be predictive of poor outcome at 5 years for breast cancer patients. In this case a group of patients relapsing at 5 years was compared in terms of gene expression to a group of patients not relapsing within 5 years.

In class discovery studies no predefined groups are available and the interest is in findings new groupings, usually called bioprofiles, using the available expression measures.

The standard statistical method to perform class discovery is cluster analysis that received great expansion due to gene expression studies. It is worth saying that cluster analysis is a powerful yet tricky method that should be applied taking care of outliers, stability of results, number of suspected profiles, and so on. These aspects are very hard to face with thousands of transcript to be analyzed. Even more subtle is the problem of the interpretation of the clusters obtained in terms of disease profiles and the definition of a rule to define the discovered profiles.

Alternatively, classical multivariate methods, such as principal components analysis (PCA), are gaining relevance for visualization of high-dimensional data (Fig. 32.2) through data reduction [32.38, 39]. Recent work on lung cancer highlighted different patterns of

Fig. 32.2 PCA-based biplot for evaluating association between genes involved in cell-polarity pathway and a set of tumor cell-lines from the NCI60 panel. Expression values of 76 genes as profiled on 23 tumor cell lines (6 colon, 6 breast, 8 renal and 3 leukemia) were obtained from the well known NCI60 dataset [32.44,45], publicly available online on the ArrayExpress website with accession number E-GEOD-5720 (hybridization performed with human HG-U133A/133B Affymetrix GeneChip). To interpret results, let us use some examples. Cell lines labeled with *olive green points* (colon and ER+ breast cancer) are positively associated with the set of genes on the lower part of the *right vertical axis*, namely genes from CLDN15 and CDH1; on the other hand, cell lines labeled with *light blue squares* (renal lines and ER- breast cancer) are negatively associated with this set of genes, but show some mild positive association with genes on the *left side of the upper horizontal axis*, namely genes from CDH2 to ZYX

expression between healthy subjects and mesotelioma-affected subjects using multivariate visualization techniques and evaluated relationships between gene-sets involved in different biological pathways via partial least squares regression [32.40, 41]. It is relevant to note that all of these methods will have to face relevant challenges, in the perspective of being applied in the context of NGS data, for which a considerable amount of issues has already arisen [32.42]. The use of multivariate techniques is thus a principled way for making it through the steps of data dimensionality reduction, data normalization, and integration, which can lead to higher chances of driving the exploration of distinct bioprofiles [32.43].

The work of *Perou* and colleagues [32.5] is an important example of class discovery by cluster analysis in a major pathology such as breast cancer. In their work, the authors found genes distinguished between estrogen positive cancer with luminal characteristics and estrogen negative cancers. Among these two subgroups, one had a basal characterization and the other showed patterns of up-regulation for genes linked to oncogeneErb-B2. Repeated application of cluster anal-

ysis in different case series resulted in very similar groupings.

Notwithstanding the above-mentioned issues connected to cluster analysis, one of the major breakthroughs of genomic studies was actually believed to be the definition of genomic signatures/profiles by the repeated application of cluster analysis to different case series without the definition of a formal rule for class assignment of new subjects.

Profiles may then be correlated with clinical outcome as was done for breast cancer by *van't Veer* and colleagues [32.37]. Now, more than 10 years after this study, it is not yet clear which is the real contribution of microarray-based gene expression profiling to breast cancer prognosis. Of all the so-called *first-generation* signatures, only oncotype DX [32.46], a qRT-PCR based analysis of 21 genes, has reached level II of evidence to support tumor prognosis and has been included in the National Comprehensive Cancer Network guidelines, whereas the remaining signatures have only obtained level III of evidence so far [32.47]. Reasons for this are, among the others, a lack of stabil-

ity in terms of genes that the lists are composed of and strong time-dependence, i. e., reduced prognostic value after 5 to 20 years of follow-up.

Another, and more important, issue for prognostic/prediction studies is connected to the design of the study itself. In fact, a prognostic study should be planned by defining a cohort that will be followed during time while a case control study may be only suggestive of signatures to be considered for clinical management. See letter [32.48] commenting reference [32.49] for concerns regarding such an issue.

32.2.1 Multiple Comparisons and the False Discovery Rate

Class comparison in genomic-wide studies is one of the most common and challenging applications since the advent of microarray technology. The first study on predictive signatures in breast cancer in 2002 [32.6] was mainly a class comparison study.

From the statistical viewpoint one of the first problems evidenced was the large number of statistical tests performed in such an analysis. In particular, the classical control for false positives, emphasizing the specificity of the screening appeared from the beginning to be too restrictive with the cost of false negatives too high.

To understand such an issue, let us suppose to have to compare the gene expression in a group of tumor tissues with that of a group of normal tissues. For each gene a statistical test controlling the probability of saying that a gene is different when in fact it is not (false positive, FP), is performed. Such an error is called type one error and its level is generally called α and fixed at a 5% level. The problem is that a test at α level is performed for each gene. Therefore, if the probability of making a mistake (FP) is 0.05, while the probability of not making a mistake is 0.95 (this is the probability of saying the gene is not differentially expressed when it is not, true negative), when performing, say, 1000 tests the probability of not making any mistake is 0.95^{1000}, which is practically 0. Accordingly, the probability of at least one FP is practically 1. How can the specificity of the experiment be controlled? A large number of procedures is available, the most simple and known is the Bonferroni correction. Let us see how it works.

In particular, if n tests are performed at α level, the probability of not having any false positive is $(1 - \alpha)^n$, therefore the probability of making at least one false positive is $1 - (1 - \alpha)^n$, which can be approximated as $1 - n\alpha$ (for small α). The Bonferroni correction originates from this. In fact, if the tests are performed at

level $\alpha_{\mathrm{B}} = \alpha/n$, then we can expect to have no false positive among the genes declared differentially expressed at α level. This is, in fact, at the cost of a large number of false positives. In genomic experiments, when thousands of tests are performed, the Bonferroni significance level is so low that very few genes can easily pass the first screening probably paying too high costs in terms of genes declared not significantly differentially expressed when actually they have a differential expression. The balance between specificity and sensibility is a fairly old problem in screening problems, which is exacerbated with high-throughput data analysis.

One of the most common approaches applied in such a context is the proposal of *Benjamini* and *Hochberg* [32.50] called the false discovery rate (FDR) trying to control the number of false positives among the genes declared significant.

To better understand the novelty of FDR, let us suppose to have M genes to be considered in the high-throughput experiment, N of the M genes are truly differentially expressed while P are not. Performing the appropriate statistical test NS of the M genes are declared not different between groups under comparison while S are significantly different (Fig. 32.1). The type one error rate α (FPR) controls the number of FP with respect to N, while using the Bonferroni correction the probability that FP is greater than 0 is controlled. The FDR changes perspective and considers the columns of the table instead of the rows. FDR controls the number of FP with respect to S. If, for example, 10 genes are declared differentially expressed with an FDR of 20%, it is expected that 2 be false positives. This may allow greater flexibility in the managing of the screening phase of the analysis (see Fig. 32.3 for a graphical representation of results from a class comparison microarray study, with an application of FDR concepts). The problem first solved by *Benjamini* and *Hochberg* was basically how to estimate FDR and different proposals have appeared since then, for example the q-value of *Storey* [32.51].

In general, omic and multiplexed diagnostic technologies with the ability to produce vast amounts of biomolecular data, have vastly outstripped our ability to sensibly deal with this data deluge and extract useful and meaningful information for decision making. The producers of novel biomarkers assume that an integrated bioinformatics and biostatistics infrastructure exists to support the development and evaluation of multiple assays and their translation to the clinic. Actually, the best scientific practice for the use of high-throughput data is still to be developed. In this perspective, the existence of
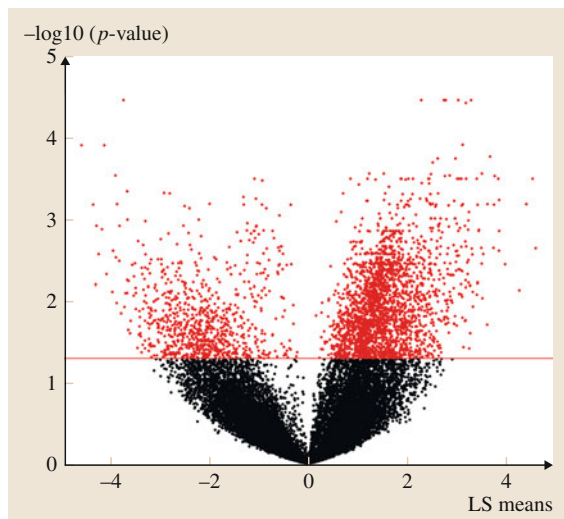
Fig. 32.3 Volcano plot of differential gene expression pattern between experimental groups. On the *x-axis* the least squares (LS) means (i. e., difference of mean expression on $\log_2$ scale between experimental groups) and on the *y-axis* $-\log_{10}$ transformed *p*-values corrected for multiplicity using the FDR method from *Benjamini* et al. [32.50] are reported. The *horizontal red line* corresponds to a cut-off for the significance level α at 0.05. *Points* above this threshold represent genes which are actually differentially expressed between experimental groups, and that are to be further investigated

advanced computational technologies for bioinformatics is irrelevant along the translational research process unless supporting biostatistical evaluation infrastructures exist to take advantage of developments in any technology.

In this sense, a key problem is the fragmentation of quantitative research efforts. The analysis of high dimensional data is mainly conducted by researchers with limited biostatistical experience using standard software without the knowledge of the underlying statistical principles of the methodology then exposing the results to a wide uncertainty not only due to sample size limitations. Moreover, so far, a large amount of biostatistical methods and software tools supporting bioinformatics analysis of genomic/proteomic data has been provided but reference standardized analysis procedures coping with suitable preprocessing and quality control approaches on raw data coming from omic and multiplex assays are still waiting for development. Formal initiatives for the integration of biostatistical research groups with functional genomics and proteomics labs are one of the major challenges in this context. In fact, besides the development of innovative biostatistics and bioinformatics tools, a major key of success lies in the ability to integrate different competencies. Such an integration cannot be simply demanded for the development of software, such as the ArrayTrack initiative, but needs to develop integrated skills assisted by a software platform able to outline the analysis plan. In such a context, different strategies can be adopted from open software, such as R and bioconductor, to commercial ones such as SAS/JMP genomics.

In a functional dynamic perspective, to the characterization of the bio-profiles of cancer affected patients, is added the complexity related to the prolonged follow-up of patients with the necessity of the registration of the event-history of possible adverse events (local recurrence and/or metastasis) before death, that may offer useful insight into disease dynamics to identify a subset of patients with worse prognosis and better response to the therapy. This makes it necessary to develop strategies for the integration of clinical and follow-up information with those deriving from genetic and molecular characterizations.

The evaluation and benchmarking of new analytical processes for the discovery, development, and clinical validation of new diagnostic/prognostic biomarkers is an extremely important problem especially in a fast growing area such as translational research based on functional genomics/proteomics. In fact, the presentation of overoptimistic results based on the unsuited application of biostatistical procedures can mask the true performance of new biomarker/bioprofiles and create false expectations about its effectiveness.

Guidelines for omic and cross-omic studies should be defined through the integration of different competencies coming from clinical-translational, bioinformatics, and biostatistics research competencies.

This integrated contribution from multidisciplinary research teams will have a major impact on the development of standard procedures that will standardize the results and make research more consistent and accurate according to relevant bioanalytical and clinical targets.

Genes dataset

Feature selection — **Multiple test, *q*-value** — **Gene clustering**

Feature extraction — **Principal component analysis**

Classification — **Logistic model**

Performance assessment — **(Cross-validated) error rate**
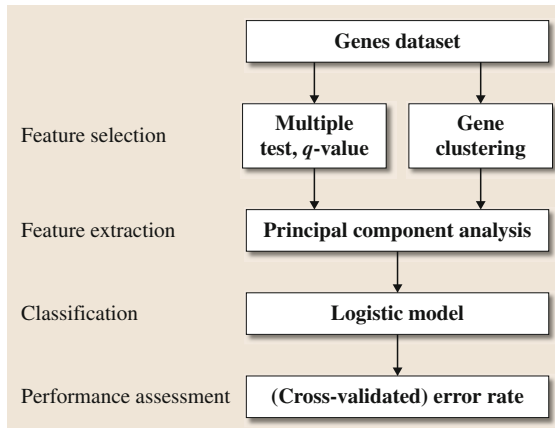
Fig. 32.4 Microarray studies have provided insight on global gene expression in cells and tissues with the expectation of prognostic assessments improvement. The identification of genes whose expression levels are associated with recurrence might also help better discriminating those subjects who are likely to respond to the various tailored systemic treatments. However, microarray experiments raised several questions to the statistical community about the design of the experiments, data acquisition and normalization, supervised and unsupervised analysis. All these issues are burdened by the fact that typically the number of genes being investigated far exceeds the number of patients. It is well-recognized that too large a number of predictor variables affects the performance of classification models: Bellman coined the term *curse of dimensionality* [32.52], referring to the fact that in the absence of simplifying assumptions, the sample size needed to estimate a function of several variables to a given degree of accuracy (i. e., to get a reasonably low-variance estimate) grows exponentially with the number of variables. To avoid this problem, feature selection and extraction issue play a crucial role in microarray analysis. This has led several researchers to find it judicious to filter out genes that do not change their expression level *significantly*, reducing the complexity of the data and improving the signal to noise ratio. However, the adopted *measure of significance* in filtering (the implicitly controlled error measure) is not often easy to interpret in terms of the simultaneous testing of thousands of genes. Moreover, gene expressions are usually filtered on a per-gene basis and seldom taking into account the correlation between different gene expressions. This filtering approach is commonly used in most current high-throughput experiments whose main objective is to detect differentially expressed genes (active genes) and, therefore, to generate hypotheses rather than to confirm them. All these methods, based on a *measure of significance*, select genes from a supervised perspective, i. e., accounting for the outcome of interest (the subject status). However, an unsupervised approach might be useful in order to reveal the pattern of associations among different genes making it possible to single out redundant information. The figure shows the data analysis pipeline developed in many papers dealing with expressions from high throughput experiments

Integration and standardization of approaches for assessment of diagnostic and prognostic performance is a key issue. Many of the clinical and translational research groups have chosen different approaches for biodata modeling, tailored to specific types of medical data. However, very few proper benchmarking studies of algorithm classes have been performed worldwide and fewer examples of best practice guidelines have been produced. Similarly, few studies have closely examined the criteria under which medical decisions are made. The integrating aspects of this theme relates to methods and approaches for inference, diagnosis, prognosis, and general decision making in the presence of heterogeneous and uncertain data.

A further priority is to ensure that research in biomarker analysis is designed and informed from the outset to integrate well with clinical practice (to facilitate widespread clinical acceptance) and that it exploits cross-over between methods and knowledge from different areas (to avoid duplication of efforts to facilitate rapid adoption of good practice in the development of this healthcare technology). Reference problems are related to the assessment of improved diagnostic and prognostic tools in the clinical setting, resorting to observational and experimental clinical studies from phase I to phase IV and the integration with studies on therapy efficacy which would involve bioprofile and biopattern analysis.

In this perspective, the integration of different omic data is a well-known issue that is receiving increasing attention in biomedical research [32.53, 54], and which questions the capability of researchers to make sense out of a huge amount of data with very different features. Since this integration can not only be seen as an IT problem, proper biostatistical approaches need to be taken into account that consider the multivariate nature of the problem in the light of exploiting the maximum prior information about the biological patterns underlying the clinical problem.

A critical review of microarray studies was performed earlier in a paper by *Dupuy* and *Simon* [32.10], in which a thorough analysis of the major limitations and pitfalls of 90 microarray studies published in 2004 concerning cancer outcome was done (see Fig. 32.4 for a general pipeline for high-throughput experiments). Integrated into this review was the attempt to write guidelines for statistical analysis and reporting of gene expression microarray studies. Starting from this work, it will be possible to extend the outlined criticisms to a wider range of omic studies, in order to produce updated guidelines useful for biomolecular researchers.

In the perspective of integrating omic data coming from different technologies, a comparison of microarray data with NGS platforms will be a relevant point [32.55–57]. Due to the lack of sufficiently standardized procedures for processing and analyzing NGS data, much attention will be given to the process of data generation and of quality control evaluation. Such an integration is crucial because, though capabilities of NGS platforms mostly outperform those of microarrays, protocols of management and data analysis are typically very time-consuming, thus making it impractical to be used for in-depth analysis of large samples.

Of note, one of the ultimate goals of biomedical research is to connect diseases to genes that specify their clinical features and to drugs capable of treating them. DNA microarrays have been used for investigating genome-wide expression of common diseases producing a multitude of gene signatures predicting survival, whose accuracy, reproducibility, and clinical relevance has, however, been debated [32.48, 49, 58, 59]. Moreover, the regulatory relationships between the signature genes have rarely been investigated, largely limiting their biological understanding. The genes, indeed, never act independently from each other. Rather, they form functional connections that coordinate their activity. Hence, it is fundamental that in each cell in every life stage, regulatory events take place in order to keep the healthy steady state. Any perturbation of a gene network, in fact, has a dramatic effect on our life, leading to disease and even death.

32.3 Nanotechnology in Human Healthcare

The prefix *nano* is from the Greek world meaning *dwarf*. Nanotechnology refers to the science of materials whose functional organization is on the nanometer scale, that is 10^{-9} m. Starting from ideas originating in physics in the 1960s and boosted by the need of miniaturization (i.e., speed) of the electronic industry the field has grown rapidly. Today, nanotechnology is gaining an important place in the medicine of the future. In particular, by using the patho-physiological conditions of diseased and inflamed tissues it is possible to target nanoparticles and with them drugs, genes, and diagnostic tools.

Moreover, the spatial and/or temporal contiguity of data from NGS and nanobiotech diagnostic approaches imposes the adoption of methods related to signal analysis which are still to be introduced in standard software, being related to statistical functional data analysis methods. Therefore, the extension of the multivariate statistical methodologies adopted so far is requested in a functional data context; a problem that has already been met in the analysis of mass spectrometry data from proteomic analyses.

Nanotechnology-based platforms for the high-throughput, multiplexed detection of proteins and nucleic acids actually promise to bring substantial advances in molecular diagnostics. Forecasted applications of nano-diagnostic devices are related to the assessment of the dynamics of cell process for a deeper knowledge of the ongoing etio-pathological process at the organ, tissue, and even single cell level.

NGS is a growing revolution in genomic nanobiotechnologies that parallelized the assay process, integrating reactions at the micro or nano scale on chip surfaces, producing thousands or millions of sequences at once. These technologies are intended to lower the costs of nucleic acid sequencing far beyond that possible with earlier methods.

Concerning cancer, a key issue is related to the improvement of early detection and prevention through the understanding of the cellular and molecular pathways

of carcinogenesis. In such a way it would be possible to identify the conditions that are precursors of cancer before the start of the pathological process, unraveling its molecular origins. This should represent the next frontiers of bioprofiling to allow the strict monitoring and possible reversal of the neoplastic transformation through personalized preventive strategies. Advances in nanobiotechnology enables the visualization of changes in tissues and physiological process with a subcellular real-time spatial resolution. This is a revolution that can be compared to the *daguerreotype* pictures from current high-throughput multiplex approaches to the digital high resolution of next generation diagnostic devices.

Enormous challenges remain in managing and analyzing the large amounts of data produced. Such evolution is expected to have a strong impact in terms of personalized medical prevention and treatment with considerable effects on society. Therefore, the success will be strongly related to the capability of integrating data from multiple sources in a robust and sustainable research perspective, which could enhance the transfer of high-throughput molecular results to novel diagnostic and therapy application.

The new framework of nanobiotechnology approaches in biomedical decision support according to improved clinical investigation and diagnostic tools is emerging. There is a general need for guidelines for biostatistics and bioinformatics practice in the clinical translation and evaluation of new biomarkers from cross-omic studies based on hybridization, NGS, and high-throughput multiplexed nanobiotechnology assays. Specifically, the major topics concern: bioprofile discovery, outcome analysis in the presence of complex follow-up data, assessment of diagnostic, and prognostic values of new biomarkers/bioprofiles.

32.3.1 New Nanobiotechnological Perspectives in Diagnosis

Current molecular diagnostic technologies are not conceived to manage biological heterogeneity in tissue samples, in part because they require homogeneous preparation, leading to a loss of valuable spatial information regarding the cellular environment and tissue morphology. The development of nanotechnology has provided new opportunities for integrating morphological and molecular information and for the study of the association between observed molecular and cellular changes with clinical-epidemiological data.

Concerning specific approaches, bioconjugated quantum dots (QDs) [32.60–63] have been used to quantify multiple biomarkers in intact cancer cells and tissue specimens, allowing the integration of traditional histopathology versus molecular profiles for the same tissue [32.64–69]. Current interest is focused on the development of nanoparticles with one or multiple functionalities. For example, binary nanoparticles with two functionalities have been developed for molecular imaging and targeted therapy. Bioconjugated QDs, which have both targeting and imaging functions, can be used for targeted tumor imaging and for molecular profiling applications.

Nanoparticle material properties can be exploited to elicit clinical advantage for many applications, such as for medical imaging and diagnostic procedures. Iron oxide constructs and colloidal gold nanoparticles can provide enhanced contrast for magnetic resonance imaging (MRI) and computed tomography (CT) imaging, respectively [32.70, 71]. QDs provide a plausible solution to the problems of optical in vivo imaging due to the tunable emission spectra in the near-infrared region, where light can easily penetrate through the body without harm and their inherent ability to resist bleaching [32.72]. For ultrasound imaging, contrast relies on impedance mismatch presented by materials that are more rigid or flexible than the surrounding tissue, such as metals, ceramics, or microbubbles [32.73]. Continued advancements of these nano-based contrast agents will allow clinicians to image the tumor environment with enhanced resolution for a deeper understanding of disease progression and tumor location.

Additional nanotechnologically-based detection and therapeutic devices have been made possible using photolithography and nucleic acid chemistry [32.74, 75]. The same technology that enabled integrated circuitry, produced microelectromechanical systems (MEMS) for selective molecular sensing, sieving, and controlled drug release [32.76]. Microfluidic systems, also known as *lab-on-chip*, are fabricated by soft lithography of inexpensive polymers [32.77]. Micro- and nanoarrays, have experienced success for molecular diagnostic, genotyping, and biomarker-guided therapeutic targeting [32.76, 78, 79]. Moreover, advances in proteomics have been made possible due to the technical refinement of lithographic resolution [32.80]. Recent interest in nanowires [32.81, 82] and cantilever arrays [32.83–85] for biomarker detection has shown promise. The former are biologically gated transistors able to detect multiple, real-time, simultaneous molecular events. These innovative nanodevices equal or exceed the sensitivity of commercially available approaches [32.86] and are anticipated to be clinically available in the near future.

Highly sensitive biosensors that recognize genetic alterations or detect molecular biomarkers at extremely low concentration levels are crucial for the early detection of diseases and for early stage prognosis and therapy response. Nanowires have been used to detect several biomolecular targets such as DNA and proteins [32.82, 87]. The identification of DNA alterations is crucial to better understand the mechanism of a disease such as cancer and to detect potential genomic markers for diagnosis and prognosis. Other studies have reported the development of a three-dimensional gold nanowire platform for the detection of mRNA with enhanced sensitivity from cellular and clinical samples. Highly sensitive electrochemical sensing systems use peptide nucleic acid probes to directly detect specific mRNA molecules without PCR amplification steps [32.88–90].

The development of immuno and aptamer-based nanowire biosensors to detect cancer biomarkers such as VEGF [32.91] and CA125 [32.92], or SARS virus N-protein [32.93] has shown a great sensitivity, providing the potential use of these nanodevices for point-of-care diagnostic applications. To improve the diagnostic efficacy of the biosensors, a multiplexed approach is needed to accurately identify heterogeneous diseases such as cancer [32.94].

Cantilever nanosensors have also been used to detect minute amount of protein biomarkers. Label-free resonant microcantilever systems have been developed to detect the ng/mL level of alpha-fetoprotein, a potential marker of hepatocarcinoma, providing an opportunity for early disease diagnosis and prognosis [32.95]. Nanofabricated and functionalized devices such as nanowires and nanocantilevers are fast, multiplexed, and label-free methods that provide extraordinary potential for the future of personalized medicine.

The combination of data from multiple imaging techniques offers many advantages over data collected from a single modality. Potential advantages include: improved sensitivity and specificity of disease detection and monitoring, smarter therapy selection based on larger data sets, and faster assessment of treatment efficacy. The successful combination of imaging modalities, however, will be difficult to achieve with multiple contrast agents. Multimodal contrast agents stand to fill this niche by providing spatial, temporal, and/or functional information that corresponds with anatomic features of interest.

There is also great interest in the design of multifunctional nanoparticles, such as those that combine contrast and therapeutic agents. The integration of diagnostics and therapeutics, known as *theranostics*, is attractive because it allows the imaging of therapeutic delivery, as well as follow-up studies to assess treatment efficacy.

Finally, a key direction of research is the optimization of biomarker panels via principled biostatistics approaches for the quantitative analysis of molecular profiles for clinical outcome and treatment response prediction. The key issues that will need to be addressed are: (i) a panel of tumor markers will allow more accurate statistical modeling of the disease behavior than relying on single tumor markers; and (ii) the combination of tumor gene expression data and molecular information of the cancer microenvironment is necessary to define aggressive phenotypes of cancer, as well as for determining the response of early stage disease to treatment (chemotherapy, radiation, or surgery).

Currently, the major tasks in biomedical nanotechnology are (i) to understand how nanoparticles interact with blood, cells, and organs under in vivo physiological conditions and (ii) to overcome one of their inherent limitations, that is, their delivery to diseased sites or organs [32.96–98]. Another major challenge is to generate critical studies that can clearly link biomarkers with disease behaviors, such as the rate of tumor progression and different responses to surgery, radiation or drug therapy [32.99]. The current challenge is, therefore, related to the advancement of biostatistics and biocomputing techniques for the analysis of novel high-throughput biomarkers coming from nanotechnology applications. Current applications involve high-throughput analysis of gene expression data and for multiplexed molecular profiling of intact cells and tissue specimens. The advent of fast and low cost high-throughput diagnostic devices based on NGS approaches appears to be of critical relevance for improving the technology transfer to disease prevention and clinical strategies.

The development of nanomaterials and nanodevices offers new opportunities to improve molecular diagnosis, increasing our ability to discover and identify minute alterations in DNA, RNA, proteins, or other biomolecules. Higher sensitivity and selectivity of nanotechnology-based detection methods will permit the recognition of trace amounts of biomarkers which will open extraordinary opportunities for systems biology analysis and integration to elicit effective early detection of diseases and improved therapeutic outcomes; hence paving the way to achieving individualized medicine.

Effective personalized medicine depends on the integration of biotechnology, nanotechnology, and informatics. Bioinformatics and nanobioinformatics are cohesive forces that will bind these technologies together. Nanobioinformatics represents the application of information science and technology for the purpose of research, design, modeling, simulation, communication, collaboration, and development of nano-enabled products for the benefit of mankind. Within this framework a critical role is played by evaluation and benchmarking approaches according to a robust Health Technology Assessment approach; moreover the development of enhanced data analysis approaches for the integration of multimodal molecular and clinical data should be based on up to date and validated biostatisical approaches. Therefore, in the developing nanobiotechnology era, the role of biostatistical support to bioinformatics is definitely essential to prevent loss of money and suboptimal developments of biomarkers and diagnostic disease signature approaches of the past, which followed a limited assessment according to a strict business perspective rather than to social sustainability.

32.4 Discussion

Concerning the relevance and impact for national health systems, it is forecasted that current omic approaches based on nanobiotechnology will contribute to the identification of next generation diagnostic tests which could be focused on primary to advanced disease prevention by early diagnosis of genetic risk patterns, or the start or natural history of the pathological process of multifactor chronic disease by the multiplexed assessment of both direct and indirect, inner genetic, or environment causal factors.

A benefit of such a development would be finally related to the reduction of costs in the diagnostic process since nanobiotechological approaches seem best suited in the perspective of points-of-care POC diagnostic facilities which could be disseminated in large territories with a reduced number of excellence clinical facilities with reference diagnostic protocols. Nanomaterials are providing the small, disposable lab-on-chip tests that are leading this new approach to healthcare. A variety of factors are provoking calls for changes in how diagnosis is managed. The lack of infrastructure in the developing world can be added to the inefficiency and cost of many diagnostic procedures done in central labs, rather than by a local doctor. For the developed world, an increasingly elderly population is going to exacerbate demand on healthcare and any time-saving solutions will help deal with this new trend. POC devices are looking to reduce the dependence on lab tests and make diagnosis easier, cheaper, and more accessible for countries lacking healthcare infrastructure.

A key role in the overall framework will be played by data analysis under principled biostatistical approaches to develop suitable guidelines for data quality analysis, the following extraction of relevant information and communication of the results in an ethical and sustainable perspective for the individual and society.

The proper, safe and secure management of personalized data in a robust and shared bioethical reference framework is, indeed, expected to reduce the social costs related to unsuited medicalization through renewed preventive strategies. A strong biostatistical based Health Technology Assessment phase will be essential to avoid the forecasted drawbacks of the introduction of such a revolution in prevention and medicine.

To be relevant for national health services, research on biostatistics and bioinformatics applied to nano-biotechnology should exploit its transversal role across multiple applied translational research projects on biomarker discovery, development, and clinical validation until their release for routine application for diagnostic/prognostic aims. Objectives that would enable an accelerated framework for translational research since the involvement of quantitative support are listed here:

- Technological platforms for the developments in the fields of new diagnostic prevention and therapeutic tools. In the context of preventing and treating diseases, the objectives are to foster academic and industrial collaboration through technological platforms where multidisciplinary approaches using cutting edge technologies arising from genomic research may contribute to better healthcare and cost reduction through more precise diagnosis, individualized treatment, and more efficient development pathways for new drugs and therapies (such as the selection of new drug candidates), and other novel products of the new technologies.
- Patentable products: customized array and multiplex design with internal and external controls for optimized normalization. Validation by double checked expression results for genes or protein in the customized array and multiplex assays.

Patenting of validated tailor-made cDNA/proteomic arrays that encapsulate gene/protein signatures related to the response to the therapy with optimized cost/effectiveness properties.

A robust, multidisciplinary quantitative assessment framework in translational research is a global need, which should characterize any specific laboratory and clinical translation project. However, the quantitative assessment phase is rarely based on an efficient cooperation between biologists, biotechnologists, and clinicians with biostatisticians, with relevant skills in this field. This represents a major limitation to the rapid transferability of basic research results to healthcare. Such a condition is solved in the context of pharmacology in the research and development of new drugs to their assessment in clinical trials, whereas, for diagnostic/prognostic biomarkers, this framework is still to be fully defined.

Such a gap is wasting resources and is malpractice in the use of biomarkers and related bioprofiles for clinical decision making in critical phases of chronic and acute major diseases like cancer and cardiovascular pathologies.

References

32.1 A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, C. Smigal, M.J. Thun: Cancer statistics, CA – Cancer J. Clin. **56**, 106–130 (2006)

32.2 G. Arpino, R. Laucirica, R.M. Elledge: Premalignant and in situ breast disease: Biology and clinical implications, Ann. Intern. Med. **143**, 446–457 (2005)

32.3 B. Vogelstein, E.R. Fearon, S.R. Hamilton: Genetic alteration during colorectal tumor development, N. Engl. J. Med. **319**, 525–532 (1988)

32.4 D. Hanahan, R.A. Weinberg: The hallmarks of cancer, Cell **100**, 57–70 (2000)

32.5 C.M. Perou, T. Sørlie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, C.A. Rees, J.R. Pollack, D.T. Ross, H. Johnsen, L.A. Akslen, Ø. Fluge, A. Pergamenschikov, C. Williams, S.X. Zhu, P.E. Lønning, A. Børresen-Dale, P.O. Brown, D. Botstein: Molecular portraits of human breast tumours, Nature **406**, 747–752 (2000)

32.6 M.J. van de Vijver, Y.D. He, L.J. van't Veer, H. Dai, A.A.M. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E.T. Rutgers, S.H. Friend, R. Bernards: A gene-expression signature as a predictor of survival in breast cancer, N. Engl. J. Med. **347**, 1999–2009 (2002)

32.7 Y. Wang, J.G.M. Klijn, Y. Zhang, A.M. Sieuwerts, M.P. Look, F. Yang, D. Talantov, M. Timmermans, M.E. Meijer-van Gelder, J. Yu, T. Jatkoe, E.M.J.J. Berns, D. Atkins, J.A. Foekens: Gene expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer, The Lancet **365**, 671–679 (2005)

32.8 X. Ma, R. Salunga, J.T. Tuggle, J. Gaudet, E. Enright, P. McQuary, T. Payette, M. Pistone, K. Stecker, B.M. Zhang, Y. Zhou, H. Varnholt, B. Smith, M. Gadd, E. Chatfield, J. Kessler, T.M. Baer, M.G. Erlander, D.C. Sgroi: Gene expression profiles of human breast cancer progression, Proc. Natl. Acad. Sci. USA **100**, 5974–5979 (2003)

32.9 MAQC Consortium: The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements, Nat. Biotechnol. **24**, 1151–1161 (2006)

32.10 A. Dupuy, R.M. Simon: Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting, J. Natl. Cancer Inst. **99**, 147–157 (2007)

32.11 A.H. Sims, K.R. Ong, R.B. Clarke, A. Howell: High-throughput genomic technology in research and clinical management of breast cancer. Exploiting the potential of gene expression profiling: Is it ready for the clinic?, Breast Cancer Res. **8**, 214 (2006)

32.12 W. Boecker, H. Buerger, K. Schmitz, I.A. Ellis, P.J. van Diest, H.P. Sinn, J. Geradts, R. Diallo, C. Poremba, H. Herbst: Ductal epithelial proliferations of the breast: A biologic continuum? Comparative genomic hybridization and high-molecular-weight cytokeratin expression patterns, J. Pathol. **195**, 415–421 (2001)

32.13 J.G. Herman, S.B. Baylin: Gene silencing in cancer in association with promoter hypermethylation, N. Engl. J. Med. **349**, 2042–2054 (2003)

32.14 K.Y. Tai, S.G. Shiah, Y.S. Shieh, Y.R. Kao, C.Y. Chi, E. Huang, H.S. Lee, L.C. Chang, P.C. Yang, C.W. Wu: DNA methylation and histone modification regulate silencing of epithelial cell adhesion molecule for tumor invasion and progression, Oncogene **26**, 3989–3997 (2007)

32.15 S.L. Berger: Histone modifications in transcriptional regulation, Curr. Opin. Genet. Dev. **12**, 142–148 (2002)

32.16 M.V. Iorio, M. Ferracin, C.G. Liu, A. Veronese, R. Spizzo, S. Sabbioni, E. Magri, M. Pedriali, M. Fabbri, M. Campiglio, S. Ménard, J.P. Palazzo, A. Rosenberg, P. Musiani, S. Volinia, I. Nenci, G.A. Calin, P. Querzoli, M. Negrini, C.M. Croce: MicroRNA gene expression deregulation in human breast cancer, Cancer Res. **65**, 7065–7070 (2005)

32.17   M.J. Bissell, D. Radisky: Putting tumours in context, Nat. Rev. Cancer **1**, 46–54 (2001)

32.18   M. Al-Hajj, M.S. Wicha, A. Benito-Hernandez, S.J. Morrison, M.F. Clarke: Prospective identification of tumorigenic breast cancer cells, Proc. Natl. Acad. Sci. **100**, 3983–3988 (2003)

32.19   D. Ponti, A. Costa, N. Zaffaroni, G. Pratesi, G. Petrangolini, D. Coradini, S. Pilotti, M.A. Pierotti, M.G. Daidone: Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties, Cancer Res. **65**, 5506–5511 (2005)

32.20   T. Reya, S.J. Morrison, M.F. Clarke, I.L. Weissman: Stem cells, cancer, and cancer stem cells, Nature **414**, 105–111 (2001)

32.21   J.B. Kim, R. Stein, M.J. O'Hare: Tumour-stromal interactions in breast cancer: The role of stroma in tumourigenesis, Tumour Biol. **26**, 173–185 (2005)

32.22   T.D. Tisty, P.W. Hein: Know thy neighbor: Stromal cells can contribute oncogenic signals, Curr. Opin. Genet. Dev. **11**, 54–59 (2001)

32.23   V. Montel, E.S. Mose, D. Tarin: Tumor-stromal interactions reciprocally modulate gene expression patterns during carcinogenesis and metastasis, Int. J. Cancer **119**, 251–263 (2006)

32.24   M.M. Mueller, N.E. Fusenig: Friends or foes – bipolar effects of the tumour stroma in cancer, Nat. Rev. Cancer **4**, 839–849 (2004)

32.25   J.D. Yager, N.E. Davidson: Estrogen carcinogenesis in breast cancer, N. Engl. J. Med. **354**, 270–282 (2006)

32.26   C.L. Wilson, A.H. Sims, A. Howell, C.J. Miller, R.B. Clarke: Effects of oestrogen on gene expression in epithelium and stroma of normal human breast tissue, Endocr. Relat. Cancer **13**, 617–628 (2006)

32.27   J. Russo, I.H. Russo: The role of estrogen in the initiation of breast cancer, J. Steroid Biochem. Mol. Biol. **102**, 89–96 (2006)

32.28   L.M. Coussens, Z. Werb: Inflammation and cancer, Nature **420**, 860–867 (2002)

32.29   O. Kovalchuk, V.P. Tryndyak, B. Montgomery, A. Boyko, K. Kutanzi, F. Zemp, A.R. Warbritton, J.R. Latendresse, I. Kovalchuk, F.A. Beland, I.P. Pogribny: Estrogen-induced rat breast carcinogenesis is characterized by alterations in DNA methylation, histone modifications and aberrant microRNA expression, Cell Cycle **6**, 2010–2018 (2007)

32.30   T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, Science **286**, 531–537 (1999)

32.31   B. Walsh: *Multiple Comparisons: Bonferroni Corrections and False Discovery Rates*, Lecture Notes for EEB, Vol. 581 (Univ. Arizona, Tucson 2004)

32.32   I.B. Jeffery, D.G. Higgins, A.C. Culhane: Comparison and evaluation of methods for generating differ-entially expressed gene lists from microarray data, BMC Bioinformatics **7**, 359 (2006)

32.33   D.M. Witten, R. Tibshirani: A comparison of fold-change and the *t*-statistic for microarray data analysis, Stanford Univ. Tech. Report (2007)

32.34   G.K. Smyth: Linear models and empirical Bayes methods for assessing differential expression in microarray experiments, Stat. Appl. Genet. Biol. **3**, article 3 (2004)

32.35   J. Xu, X. Cui: Robustified MANOVA with applications in detecting differentially expressed genes from oligonucleotide arrays, Bioinformatics **24**, 1056–1062 (2008)

32.36   N. Bassani, F. Ambrogi, R. Bosotti, M. Bertolotti, A. Isacchi, E. Biganzoli: Non-parametric MANOVA methods for detecting differentially expressed genes in real-time RT-PCR experiments, Computational Intelligence Methods for Bioinformatics and Biostatistics 2009 – Revised Selected Papers (2010) pp. 56–69

32.37   L.J. van't Veer, H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, S.H. Friend: Gene expression profiling predicts clinical outcome of breast cancer, Nature **415**, 530–536 (2002)

32.38   S. Chapman, P. Schenk, K. Kazan, J. Manners: Using biplots to interpret gene expression in plants, Bioinformatics **18**, 202–204 (2001)

32.39   O. Alter, P.O. Brown, D. Botstein: Singular value decomposition for genome-wide expression data processing and modeling, Proc. Natl. Acad. Sci. USA **97**, 10101–10106 (2000)

32.40   C. Casarsa, N. Bassani, F. Ambrogi, G. Zabucchi, P. Boracchi, E. Biganzoli, D. Coradini: Epithelial-to-mesenchymal transition, cell polarity and stemness-associated features in malignant pleural mesothelioma, Cancer Lett. **302**, 136–143 (2011)

32.41   N. Bassani, F. Ambrogi, D. Coradini, E. Biganzoli: Use of biplots and partial least squares regression in microarray data analysis for assessing association between genes involved in different biological pathways, Computational Intelligence Methods for Bioinformatics and Biostatistcs 2010, Lecture Notes in Bioinformatics **6685**, 123–134 (2011)

32.42   D. Ghosh, Z.S. Qin: Statistical Issues in the Analysis of ChIP-Seq and RNA-Seq Data, Genes **1**, 317–334 (2010)

32.43   P.J.G. Lisboa, A. Vellido, R. Tagliaferri, F. Napolitano, M. Ceccarelli, J.D. Martin-Guerrero, E. Biganzoli: Data mining in cancer research, IEEE Comput. Intell. Mag. **5**, 14–18 (2010)

32.44   U. Scherf, D.T. Ross, M. Waltham, L.H. Smith, J.K. Lee, L. Tanabe, K.W. Kohn, W.C. Reinhold, T.G. Myers, D.T. Andrews, D.A. Scudiero, M.B. Eisen, E.A. Sausville, Y. Pommier, D. Botstein, P.O. Brown, J.N. Weinstein: A gene expression database for the

molecular pharmacology of cancer, Nat. Genet. **24**, 236–244 (2000)

32.45 D.T. Ross, U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J.C.F. Lee, D. Lashkari, D. Shalon, T.G. Myers, J.N. Weinstein, D. Botstein, P.O. Brown: Systematic variation in gene expression patterns in human cancer cell lines, Nat. Genet. **24**, 227–234 (2000)

32.46 S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, W. Hiller, E.R. Fisher, D.L. Wickerham, J. Bryant, N. Wolmark: A multigene assay to predict recurrence of tamoxifen–treaten, node–negative breast cancer, N. Engl. J. Med. **351**, 2817–2826 (2004)

32.47 P.E. Colombo, F. Milanezi, B. Weigelt, J.S. Reis-Filho: Microarrays in the 2010s: The contribution of microarray–based gene expression profiling to breast cancer classification, prognostication and prediction, Breast Cancer Res. **13**, 212 (2011)

32.48 E. Biganzoli, N. Lama, F. Ambrogi, L. Antolini, P. Boracchi: Prediction of cancer outcome with microarrays, Lancet **365**, 1683 (2005)

32.49 S. Michiels, S. Koscielny, C. Hill: Prediction of cancer outcome with microarrays: A multiple random validation strategy, Lancet **365**, 488–492 (2005)

32.50 Y. Benjamini, Y. Hochberg: Controlling the false discovery rate: A practical and powerful approach to multiple testing, J. R. Stat. Soc. B **57**, 289–300 (1995)

32.51 J.D. Storey: A direct approach to false discovery rates, J. R. Stat. Soc. B **64**, 479–498 (2002)

32.52 R. Bellman: *Adaptive Control Processes: A Guided Tour* (Princeton University Press, New Jersey 1961)

32.53 B. Palsson, K. Zengler: The challenges of integrating multi–omic data sets, Nat. Chem. Biol. **6**, 783 (2010)

32.54 D.B. Searls: Data integration: Challenges for drug discovery, Nat. Rev. Drug Discov. **4**, 45–58 (2005)

32.55 S.W. Roh, G.C.J. Abell, K. Kim, Y. Nam, J. Bae: Comparing microarrays and next–generation sequencing technologies for microbial ecology research, Trends Biotechnol. **28**, 291–299 (2010)

32.56 J.C. Marioni, C.E. Mason, S.M. Mane, M. Stephens, Y. Gilav: RNA–seq: An assessment of technical reproducibility and comparison with gene expression arrays, Genome Res. **18**, 1509–1517 (2008)

32.57 J. Juhila, T. Sipilä, K. Icay, D. Nicorici, P. Ellonen, A. Kallio, E. Korpelainen, D. Greco, I. Hovatta: MicroRNA expression profiling reveals MiRNA families regulating specific biological pathways in mouse frontal cortex and hippocampus, PLoS ONE **6**, e21495 (2011)

32.58 E.E. Ntzani, J.P. Ioannidis: Predictive ability of DNA microarrays for cancer outcomes and correlates: An empirical assessment, Lancet **362**, 1439–1444 (2003)

32.59 D. Dunkler, S. Michiels, M. Schemper: Gene expression profiling: Does it add predictive accuracy

to clinical characteristics in cancer prognosis?, Eur. J. Cancer **43**, 745–751 (2007)

32.60 W.C.W. Chan, S.M. Nie: Quantum dot bioconjugates for ultrasensitive nonisotopic detection, Science **281**, 2016–2018 (1998)

32.61 P. Alivisatos: The use of nanocrystals in biological detection, Nat. Biotechnol. **22**, 47–52 (2004)

32.62 X. Michalet, F.F. Pinaud, L.A. Bentolila, J.M. Tsay, S. Doose, J.J. Li, G. Sundaresan, A.M. Wu, S.S. Gambhir, S. Weiss: Quantum dots for live cells, in vivo imaging, and diagnostics, Science **307**, 538–544 (2005)

32.63 X. Gao, L. Yang, J.A. Petros, F.F. Marshall, J.W. Simons, S. Nie: In–vivo molecular and cellular imaging with quantum dots, Curr. Opin. Biotechnol. **16**, 63–72 (2005)

32.64 X. Gao, S. Nie: Molecular profiling of single cells and tissue specimens with quantum dots, Trends Biotechnol. **21**, 371–373 (2003)

32.65 Y. Xing, A.M. Smith, A. Agrawal, G. Ruan, S. Nie: Molecular profiling of single cancer cells and clinical tissue specimens with semiconductor quantum dots, Int. J. Nanomed. **1**, 473–481 (2006)

32.66 Y. Xing, Q. Chaudry, C. Shen, K.Y. Kong, H.E. Zhau, L.W. Chung, J.A. Petros, R.M. O'Regan, M.V. Yezhelyev, J.W. Simons, M.D. Wang, S. Nie: Bioconjugated quantum dots for multiplexed and quantitative immunohistochemistry, Nat. Protoc. **2**, 1152–1165 (2007)

32.67 M.V. Yezhelyev, A. Al-Hajj, C. Morris, A.I. Marcus, T. Liu, M. Lewis, C. Cohen, P. Zrazhevskiy, J.W. Simons, A. Rogatko, S. Nie, X. Gao, R.M. O'Regan: In situ molecular profiling of breast cancer biomarkers with multicolor quantum dots, Adv. Mater. **19**, 3146–3151 (2007)

32.68 A.A. Ghazani, J.A. Lee, J. Klostranec, Q. Ziang, R.S. Dacosta, B.C. Wilson, M.S. Tsao, W.C. Chan: High throughput quantification of protein expression of cancer antigens in tissue microarray using quantum dot nanocrystals, Nano Lett. **6**, 2881–2886 (2006)

32.69 M.V. Yezhelyev, X. Gao, Y. Xing, A. Al-Hajj, S. Nie, R.M. O'Regan: Emerging use of nanoparticles in diagnosis and treatment of breast cancer, Lancet Oncol. **7**, 657–667 (2006)

32.70 Y.X. Wang, S.M. Hussain, G.P. Krestin: Superparamagnetic iron oxide contrast agents: Physicochemical characteristics and applications in MR imaging, Eur. Radiol. **11**, 2319–2331 (2001)

32.71 Q.Y. Cai, S.H. Kim, K.S. Choi, S.Y. Kim, S.J. Byun, K.W. Kim, S.H. Park, S.K. Juhng, K.H. Yoon: Colloidal gold nanoparticles as a blood–pool contrast agent for X–ray computed tomography in mice, Invest. Radiol. **42**, 797–806 (2007)

32.72 X. Gao, W.C. Chan, S. Nie: Quantum–dot nanocrystals for ultrasensitive biological labeling and multicolor optical encoding, J. Biomed. Opt. **7**, 532–537 (2002)

32.73   J.R. Lindner: Contrast ultrasound molecular imaging: Harnessing the power of bubbles, Cardiovasc. Res. **83**, 615–616 (2009)

32.74   S.E. Fodor: Light-directed spatially addressable parallel chemical synthesis, Sciences **251**, 767–773 (1991)

32.75   J.M. Nam, C.A. Mirkin: Bio-barcode-based DNA detection with PCR-like sensitivity, J. Am. Chem. Soc. **126**, 5932–5933 (2004)

32.76   J.T. Santini, A.C. Richards, R. Scheidt, M.J. Cima, R. Langer: Microchips as controlled drug delivery devices, Angew. Chem. Int. Edn. **39**, 2396–2407 (2000)

32.77   G.M. Whitesides, E. Ostuni, S. Takayama, X.Y. Jiang, D.E. Ingber: Soft lithography in biology and biochemistry, Annu. Rev. Biomed. Eng. **3**, 335–373 (2001)

32.78   D.A. LaVan, T. McGuire, R. Langer: Small-scale systems for in vivo drug delivery, Nat. Biotechnol. **21**, 1184–1191 (2003)

32.79   M.M. Orosco, C. Pacholski, M.J. Sailor: Real-time monitoring of enzyme activity in a mesoporous silicon double layer, Nature Nano. **4**, 255–258 (2009)

32.80   M.M. Cheng, G. Cuda, Y.L. Bunimovich, M. Gaspari, J.R. Heath, H.D. Hill, C.A. Mirkin, A.J. Nijdam, R. Terracciano, T. Thundat, M. Ferrari: Nanotechnologies for biomolecular detection and medical diagnostics, Curr. Opin. Chem. Biol. **10**, 11–19 (2006)

32.81   J.R. Heath, M.E. Phelps, L. Hood: NanoSystems biology, Mol. Imaging Biol. **5**, 312–325 (2003)

32.82   Y. Cui, Q. Wei, H. Park, C.M. Lieber: Nanowire nanosensors for highly sensitive and selective detetction of biological and chemical species, Science **293**, 1289–1292 (2001)

32.83   K.M. Hansen, H.–F. Ji, G. Wu, R. Datar, R. Cote, A. Majumdar, T. Thundat: Cantilever-based optical deflection assay for discrimination of DNA single-nucleotide mismatches, Anal. Chem. **73**, 1567–1571 (2001)

32.84   G. Wu, R.H. Datar, K.M. Hansen, T. Thundat, R.J. Cote, A. Majumdar: Bioassay of prostate-specific antigen (PSA) using microcantilevers, Nat. Biotechnol. **19**, 856–860 (2001)

32.85   K.S. Hwang, S.–M. Lee, S.K. Kim, J.H. Lee, T.S. Kim: Micro- and nanocantilever devices and systems for biomolecule detection, Annu. Rev. Anal. Chem. **2**, 77–98 (2009)

32.86   J.M. Perez, F.J. Simeone, Y. Saeki, L. Josephson, R. Weissleder: Viral-induced self-assembly of magnetic nanoparticles allows the detection of viral particles in biological media, J. Am. Chem. Soc. **125**, 10192–10193 (2003)

32.87   W.U. Wang, C. Chen, K.H. Lin, Y. Fang, C.M. Lieber: Label-free detection of small-molecule-protein interactions by using nanowaire nanosensors, Proc. Natl. Acad. Sci. USA **102**, 3208–3212 (2005)

32.88   Y.Y. Degenhardt, R. Wooster, R.W. McCombie, R. Lucito, S. Powers: High-content analysis of cancer genome DNA alterations, Curr. Opin. Genet. Dev. **18**, 68–72 (2008)

32.89   C.C. WU, F.H. Ko, Y.S. Yang, D.L. Hsia, B.S. Lee, T.S. Su: Label-free biosensing of a gene mutation using a silicon nanowire field-effect transistor, Biosens. Bioelectron. **25**, 820–825 (2009)

32.90   C.A. Pratilas, D.B. Solit: Therapeutic strategies for targeting BRAF in human cancer, Rev. Recent Clin. Trials **2**, 121–134 (2007)

32.91   H.S. Lee, K.S. Kim, C.J. Kim, S.K. Hahn, M.H. Jo: Electrical detection of VEGFs for cancer diagnoses using anti-vascular endothelial growth factor aptamer-modified Si nanowire FETs, Biosens. Bioelectron. **24**, 1801–1805 (2009)

32.92   M.A. Bangar, D.J. Shirale, W. Chen, N.V. Myung, A. Mulchandani: Single conducting polymer nanowire chemiresistive label-free immunosensor for cancer biomarker, Anal. Chem. **81**, 2168–2175 (2009)

32.93   F.N. Ishikawa, H.K. Chang, M. Curreli, H.I. Liao, C.A. Olson, P.C. Chen, R. Zhang, R.W. Roberts, R. Sun, R.J. Cote, M.E. Thompson, C. Zhou: Label-free, electrical detection of the SARS virus N-protein with nanowire biosensors utilizing antibody mimics as capture probes, ACS Nano **3**, 1219–1224 (2009)

32.94   G. Zheng, F. Patolsky, Y. Cui, W.U. Wang, C.M. Lieber: Multiplexed electrical detection of cancer markers with nanowire sensor arrays, Nat. Biotechnol. **23**, 1294–1301 (2005)

32.95   Y. Liu, X. Li, Z. Zhang, G. Zuo, Z. Cheng, H. Yu: Nanogram per milliliter-level immunologic detection of alpha-fetoprotein with integrated rotating-resonance microcantilevers for early-stage diagnosis of heptocellular carcinoma, Biomed. Microdevices **11**, 183–191 (2009)

32.96   R.K. Jain: Transport of molecules, particles, and cells in solid tumors, Annu. Rev. Biomed. Eng. **1**, 241–263 (1999)

32.97   R.K. Jain: Delivery of molecular and cellular medicine to solid tumors, Adv. Drug Del. Rev. **46**, 149–168 (2001)

32.98   R.K. Jain: The next frontier in molecular medicine: Delivery of therapeutics, Nat. Med. **4**, 655–657 (1998)

32.99   M.D. Wang, J.W. Simons, S. Nie: Biomedical nanotechnology with bioinformatics – the promise and current progress, Proc. IEEE **95**, 1386–1389 (2007)